

ACCEPTED MANUSCRIPT • OPEN ACCESS

Bearing Fault Diagnosis Based on Optimized Variational Mode Decomposition and 1-D Convolutional Neural Networks

To cite this article before publication: Qinghua Wang *et al* 2021 *Meas. Sci. Technol.* in press <https://doi.org/10.1088/1361-6501/ac0034>

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is © 2021 The Author(s). Published by IOP Publishing Ltd..

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 3.0 licence, this Accepted Manuscript is available for reuse under a CC BY 3.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <https://creativecommons.org/licenses/by/3.0>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the [article online](#) for updates and enhancements.

Bearing Fault Diagnosis Based on Optimized Variational Mode Decomposition and 1-D Convolutional Neural Networks

Qinghua Wang ¹, Chenguang Yang ¹, Hongqiang Wan ¹, Donghua Deng ² and Asoke K. Nandi FREng FInstP ^{3,*}

¹ School of Mechatronic Engineering, Xi'an Technological University, Shaanxi 710021, China; wqhhuazi@163.com, 790647723@qq.com, 413162931@qq.com.

² China Petroleum Pipeline Engineering Co. Ltd., Instrumentation and Automation Room, Langfang 065000, China; 55926392@qq.com

³ Department of Electronic and Electrical Engineering, Brunel University London, UK; asoke.nandi@brunel.ac.uk; Tel.: +44 (0)1895 266119.

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

Due to the fact that measured vibration signals from a bearing are complex and non-stationary in nature, and that impulse characteristics are always immersed in stochastic noise, it is usually difficult to diagnose fault symptoms manually. A novel hybrid fault diagnosis approach is developed for the denoising signals and fault classification in this work, which combines successfully the variational mode decomposition (VMD) and one dimensional convolutional neural network (1-D CNN). VMD is utilized to remove stochastic noise in the raw signal and to enhance the corresponding characteristics. Since the modal number and penalty parameter are very important in VMD, a particle swarm mutation optimization (PSMO) as a novel optimization method and the weighted signal difference average (WSDA) as a new fitness function are proposed to optimize the parameters of VMD. The reconstructed signals of mode components decomposed by optimized VMD are used as the input of the 1-D CNN to obtain fault diagnosis models. The performance of the proposed hybrid approach has been evaluated using the sets of experimental data of rolling bearings. The experimental results demonstrate that the VMD can eliminate signal noise and strengthen status characteristics, and the proposed hybrid approach has a superior capability for fault diagnosis from vibration signals of bearings.

Keywords: fault diagnosis, bearing, variational mode decomposition (VMD), one dimensional convolutional neural network (1-D CNN), PSMO optimization method

1. Introduction

Rolling bearings are key components of rotating machines. Once there is a serious failure, it may lead to unexpected downtime and thereby resulting in huge financial losses or safety issues. According to [1], bearing-related failures account for over 30% of all rotary machines. Therefore, the

goal of accurate and effective fault diagnosis of bearings has attracted much attention to ensure high production efficiency and to improve the safety and reliability of rotary machines in industrial process. Vibration signal analysis is the most commonly used approach due to its easy measurement and high correlation with structural dynamics [2~7]. Ahmed and Nandi have explored the combined compressive sampling

method of signals, and achieved good results in bearing fault diagnosis [8,9]. Also, they have explored the compressive sensing method to produce highly compressed measurements of bearing vibration signals [10,11]. But the measured vibration signals are complex and non-stationary in nature, and meanwhile impulse characteristics of rolling bearing are usually immersed in stochastic noise. So, there are two challenging issues in bearing fault diagnosis from vibration signals. The questions are – 1) how should one effectively eliminate noise and 2) how should one extract valid fault features.

Some signal decomposition techniques, as powerful non-linear and non-stationary signal processing tools, have been developed and immediately caught much attention in the field of fault diagnosis, such as, empirical mode decomposition (EMD), Ensemble EMD (EEMD), Local mean decomposition (LMD), Singular value decomposition (SVD), empirical wavelets transform (EWT), etc. Xu proposed a new bearing fault diagnosis method combining singular value decomposition (SVD) and the squared envelope spectrum (SES) [12]. Longqing et al. put forward a fault diagnosis method of bearing clearance of reciprocating compressor based on LMD sample entropy and SVM [13]. However, these methods still have their drawbacks and limit their application. EMD has the drawbacks of mode mixing and end effects, although it has good intrinsic locally adaptive property. Rilling et al. [14] first completely probed into these issues of EMD. EEMD, an extended version of EMD, is proposed to address the mode mixing of EMD [15] but create a large amount of computational burden [16]. EWT can alleviate the sensitivity of EMD to noise and sampling, but it cannot avoid invalid decomposition caused by the concentration of boundaries. LMD can make up for the shortcomings of EMD, but both belong to recursive mode decomposition, which is affected by mode aliasing, end effects, and sampling frequency [17].

Variational mode decomposition (VMD) as a non-recursive signal processing method has been lately proposed by Dragomiretskiy and Zosso [18] in 2014. Xin et al. [19] proposed a rolling bearing fault diagnosis method based on VMD and SVM. The energy features are extracted from the intrinsic modal components decomposed by VMD and used as the input of support vector machine (SVM) to judge the working state and fault type of the bearing. Gu et al. [20] proposed a new fault diagnosis method based on statistical characteristics such as variational pattern decomposition (VMD), SVM and variance contribution rate (VCR), energy entropy (EE), and permutation entropy (PE). Liu et al. [21] proposed a feature extraction method based on parameter optimization of VMD and sample entropy, and further used to support SVM for fault diagnosis. The performance of VMD is more powerful than EMD in tone-separation [22] and EWT in feature extraction [23], because the VMD can avoid the

cumulative error and end effects. When VMD is used to decompose signals, the number of decomposition components is always preset by experience, but it directly affects the decomposition effectivity, so do penalty parameter. In order to solve the problems of choosing the values of VMD parameters, a 1.5-dimensional diagnostic method based on the optimized VMD with genetic algorithm (GA) is proposed for fault diagnosis [24]. However, a single GA cannot handle the complexity in an efficient way. Zhang et al. [25] use the improved particle swarm optimization (PSO) algorithm to optimize the parameters K and α of VMD by using the maximum value of weighted autocorrelation function maximum (AFM) as the optimization objective function. But PSO can easily get trapped in a local optimum when solving a complex multimodal problem [26], and the choice of fitness function also affects the optimization [27]. In short, VMD can be used to eliminate the noise of the non-stationary vibration signal of bearings by decomposing signals, but the problem of how to get the optimal parameters of VMD efficiently still needs research.

The second challenging issue is the question of how one should extract feature values. It is well known that the feature extraction is a bottleneck problem and deep learning methods can avoid this problem because it integrates feature extraction and classification operations into a single machine learning body to optimize jointly the classification performances [28,29]. Convolutional neural networks (CNNs) are widely used tools for deep learning which is different from the traditional feed-forward ANN because of its three architectural properties of the visual cortex cell: local receptive regions, shared weights, and subsampling. CNNs are most frequently used with two-dimensional data, such as images. Dong et al. [30] proposed a new bearing fault diagnosis method based on multiple noise reduction of SVD and EMD and improved convolution neural network (ICNN) for bearing weak fault identification. Wang et al. [31] uses CNN and AE-DNN to detect and classify faults of MMC-HVDC system. Li et al. [32] uses CNN to classify the time-frequency samples obtained by short-time Fourier transform for bearing vibration signals. However, recently, as an alternative, a modified version of 2D CNNs, called 1D Convolutional Neural Networks (1D-CNNs), has been developed [33~35]. In many applications, 1D-CNNs are desirable and thus preferable to their 2D counterparts in dealing with 1D signals due to its reduced computational complexity and the same performance with shallower architecture. The main difference between (two-dimensional) 2-D and 1-D CNN is the usage of 1-D arrays instead of 2-D matrices for both kernels and feature maps [36]. A classifier based on 1-D CNN has the ability to process the raw signal directly and to extract the representative features more precisely, without explicit feature extraction or manual selection [37].

Therefore, in order to solve these two challenging issues in bearing fault diagnosis, a novel hybrid fault diagnosis approach is developed for the denoising of signals and fault classification in this work. This combines successfully with the variational mode decomposition (VMD) and one dimensional convolutional neural network (1-D CNN). To obtain the appropriate decomposition mode, it is necessary to find a method to optimize the parameters of VMD and design an excellent fitness function.

The main contributions of this paper are summarized as follows:

1. A novel hybrid fault diagnosis approach is proposed to solve the two challenging issues of non-stationary vibration signal from bearing.

2. A novel PSMO method is proposed. The key parameters of VMD are optimized by using the PSMO method which has the advantages of both PSO and GA. Meanwhile, a weighted signal difference average (WSDA) is proposed as the fitness function.

3. The denoising signals are reconstructed by the decomposition modes and the reconstructed signals are estimated by the index of permutation entropy (PE).

4. The effectiveness of the proposed method is verified using real bearing data. The diagnosis results are analyzed from the diagnosis accuracy, confusion matrix, and data distribution, as well as compared with other methods

The rest of this paper is arranged as follows. In section 2 we propose VMD and a novel PSMO method. 1-D CNN is introduced in section 3. A novel hybrid fault diagnosis approach is proposed in section 4. The feasibility and performance of the proposed approach are discussed in Section 5. Conclusions are drawn in section 6.

2. VMD and a novel PSMO method

In 2014, Dragomiretskiy and Zosso [18] introduced a new adaptive signal processing method called VMD, which can effectively decompose non-stationary nonlinear signals. Different from Empirical mode decomposition (EMD) and LMD, VMD can effectively solves the mode aliasing problem and has significant superior anti-noise performance and higher computational efficiency. But, values of its decomposition parameters need to be artificially set, which is prone to over-decomposition or under-decomposition phenomena. In order to overcome the problem of VMD, particle swarm optimization (PSO) and genetic algorithm (GA) have been combined in a novel way to optimize the VMD parameters.

2.1 Brief introduction of VMD

VMD can non-recursively decompose a multi-component input signal into a discrete set of quasi-orthogonal band-limited intrinsic mode functions (IMFs). Each IMF component u_k has a center frequency and a finite bandwidth,

and the corresponding constrained variation model is described as follows [18].

$$\begin{cases} \min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{j\omega_k t} \right\|^2 \right\} \\ s.t. \quad \sum_k u_k = f \end{cases} \quad (1)$$

where $u_k = \{u_1, u_2, \dots, u_k\}$ is a set of modal component functions, the sum of them is the original function f , ∂_t is the partial derivative of time t , $\omega_k = \{\omega_1, \omega_2, \dots, \omega_k\}$ is the center frequency set of the modal component, $\delta(t)$ is the unit pulse function, j is an imaginary unit, and $*$ represents the convolution operation.

In order to solve equation (1), the quadratic penalty factor α and Lagrangian multiplier $\lambda(t)$ are introduced to transform the constrained variational problem into the following unconstrained variational problem.

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = & \alpha \sum_k \left\| \partial_t \left[\left(\sigma(t) + \frac{j}{\pi t} \right) e^{-j\omega_k t} \right] \right\|_2^2 \\ & + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \quad (2)$$

In detail, the implementation process of the VMD is described as follows.

Step 1, initialize mode $\{\hat{u}_k^1\}$, central frequency $\{\hat{\omega}_k^1\}$, Lagrangian multiplier λ^1 and iterations n .

Step 2, execution cycle: $n=n+1$.

Step 3, For all $\omega \geq 0$, update u_k , ω_k and λ_k .

$$u_k^{n+1} = \arg_{u_k} \min L(\{u_{i < k}^{n+1}\}, \{u_{i \geq k}^n\}, \{\omega_i^n\}, \lambda^n) \quad (3)$$

$$\omega_k^{n+1} = \arg_{\omega_k} \min L(\{u_i^{n+1}\}, \{\omega_{i < k}^{n+1}\}, \{\omega_{i < k}^n\}, \{\omega_{i \geq k}^n\}, \lambda^n) \quad (4)$$

$$\lambda_k^{n+1} = \lambda^n + \tau(f - \sum_k u_k^{n+1}) \quad (5)$$

Step 4, Repeat steps (2) to (3), until the iteration stop condition is satisfied.

$$\sum_k \left(\|u_k^{n+1} - u_k^n\|_2^2 / \|u_k^n\|_2^2 \right) < \varepsilon \quad (6)$$

Step 5, stop the iterations and obtain the IMF components.

In order to estimate the effectiveness of VMD, permutation entropy (PE) is used as it is a measure of the complexity of a time series by capturing the order relations and extracting a probability distribution of the ordinal patterns. PE was presented by Bandit and Pompe [38] for the complexity analysis of time domain data by using the comparison of neighboring values. Yan et al. [39] used PE as one nonlinear statistical measure for status characterization of rotating machines. Their study has shown promising results that PE could effectively detect and amplify the dynamic change of vibration signals, and characterize the bearing working status under different operating conditions.

PE method also has the advantages of its simplicity, extremely fast calculation, robustness, and invariance with

respect to nonlinear monotonous transformations. PE is defined as the Shannon entropy associated to the probability distribution

$$PE = - \sum_{i=1}^{D!} \pi_i \ln \pi_i$$

where, D is the embedding dimension which controls the row numbers of the new matrix. The new matrix is formed by partitioning the one-dimensional time series which need to set the embedding time delay τ and the embedding dimension D . π_i is the frequencies associated with the i possible permutation patterns, $i = 1, \dots, D!$.

2.2 A novel PSMO method

In order to decompose accurately the vibration signal of bearings by using the VMD, to prevent the phenomenon of over-decomposition or under-decomposition, and to obtain accurate bandwidth and central frequency for each component, it is necessary to optimize two parameters of the VMD, which are the modal number k and the penalty factor α .

Evolutionary algorithms, as an effective method for solving difficult optimization problems, have gained much attention. Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) are quite similar, which means that PSO and GA change from a set of points to another set of points within an iteration with visible improvement from the previous values using some probabilistic and deterministic rules. The PSO has the advantages of strong local search ability and fast convergence speed but easily falls into a local optimum and the search path is more complicated. GA has the advantages of strong global search ability but cannot handle complexity in an efficient way. Therefore, in this paper we introduce the mutation idea of GA algorithm into PSO and propose a new hybrid algorithm of PSMO by adding mutation during the update iteration of the particle swarm optimization algorithm.

The definition is as follows: suppose that in a D -dimensional search space, there is a population of M particles $X = (X_1, X_2, X_3, \dots, X_M)$. Each of these particles has a corresponding position and velocity; for example, the position of the i th particle is $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{iD})$, and the velocity is $V_i = (v_{i1}, v_{i2}, v_{i3}, \dots, v_{iD})$. The local extremum of the particle is $P_i = (p_{i1}, p_{i2}, p_{i3}, \dots, p_{iD})$, and the global optimal value of the corresponding population is $G = (g_{i1}, g_{i2}, g_{i3}, \dots, g_{iD})$. On this basis, the idea of genetic algorithm mutation is added, and the mutation probability is q , and the random number $r_d \in [0, 1]$. If $r_d > q$, the particles do not mutate, and each particle is updated according to formula (8), and the position and speed of the next generation are updated by individual local extremum and global extremum. If $r_d \leq q$, the position after the mutation is $M_i = \text{rand}(\text{lbd}, \text{ubd})$. The specific update formula is shown in (9).

$$v_i^{n+1} = \omega v_i^n + c_1 r_d (p_i - x_i^n) + c_2 r_d (g_i - x_i^n) \quad (7)$$

$$x_i^{n+1} = x_i^n + v_i^n \quad r_d > q \quad (8)$$

$$x_i^{n+1} = M_i \quad r_d < q \quad (9)$$

where ω is the inertia weight, r_d is a random number between 0~1, c_1 and c_2 are the learning factors, which represent the learning ability of local extremum and global extremum respectively.

A fitness function is needed to optimize the parameters of VMD by using PSMO. In this paper, the weighted signal difference average, WSDA, is proposed based on the signal difference average (SDA).

$$WSDA = SDA + (K \times SDA) / \beta \quad (10)$$

where K is the modal number and β is the penalty parameter. The signal difference averaging method (SDA) calculates the difference between the signal for each data point within the signals itself. A small value of SDA indicates that the signals have high similarities whereas a large value of SDA indicates that there is a lot of information loss from the signal. The equation (11) describes SDA method where y_{IMFS} is a sum of VMFs (equation 12), y_s is an input signal, and N is the data point in the signals.

$$SDA = \frac{1}{N} \sum_{n=1}^N (y_{IMFS}(n) - y_s(n)) \quad (11)$$

$$y_{IMFS} = \sum_{i=1}^K y_{IMF[i]} \quad (12)$$

The flowchart of the proposed PSMO algorithm is shown in Figure 1.

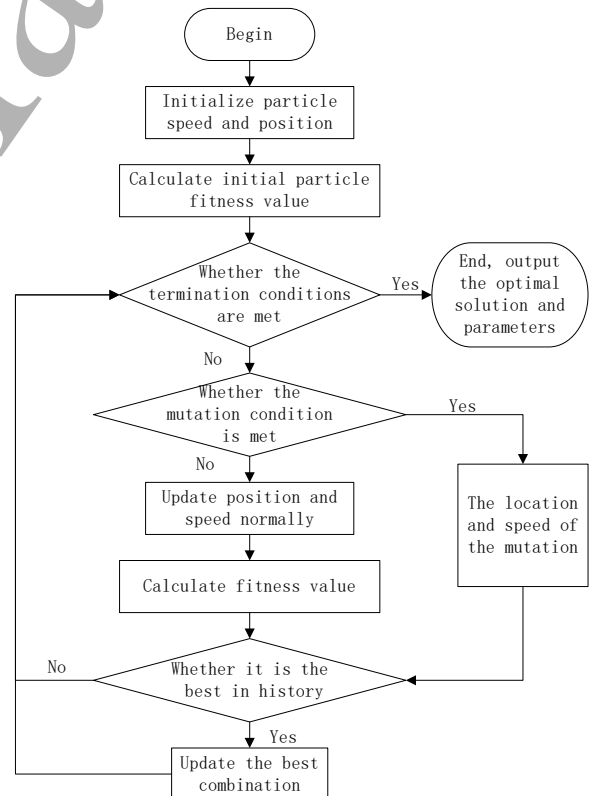


Figure 1. Flowchart of parameter optimization by the PSMO

3. 1-D Convolutional Neural Networks

More and more pattern recognition of 1-D signals uses a one-dimensional (1-D) CNN, instead of a two-dimensional (2-D) CNN, because of its reduced computing burden and often better performance. 1-D-CNN uses weight sharing, which requires fewer parameters to converge than traditional neural network models. This guarantees the convergence of 1-D-CNN earlier and faster.

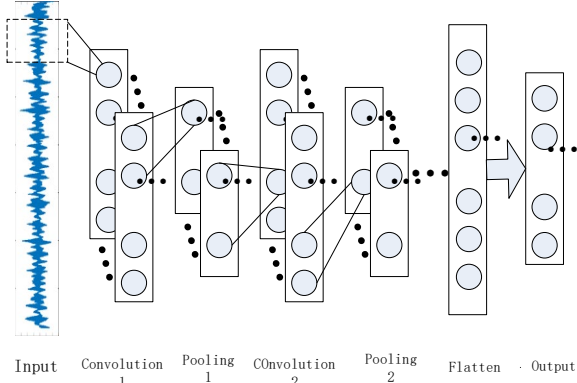


Figure 2. Architecture of 1-D CNN

The configuration of a 1-D-CNN is composed of the input layer, convolutional layer, pooling layer, full connect layer, and output layer, which are shown in Figure 2. Among these layers, there are two basic layers in CNN, which are the convolutional layer and the pooling layer. Convolution operation implements the first two properties that are local receptive regions and shared weights. The pooling operation implements the subsampling property [40]. The input layer is a passive layer that receives the raw 1-D signals and the output layer is a MLP layer with the number of neurons equal to the number of classes.

A convolutional layer consists of neurons that connect to small regions of the input and operate the convolution computation. Every kernel detects specific characteristics in any location on the input features map. The output feature map of the convolutional layer can be written as:

$$y^{l(i,j)} = K_i^l * X^{l(r^j)} \quad (13)$$

where K is the i th convolution kernel of the l th layer. $X^{l(r^j)}$ is the local area of the l th layer input, and $*$ represents the convolution operator.

Pooling layers perform down-sampling operations. Pooling functions usually include max-pooling and average-pooling.

$$p^{l(i,j)} = \frac{1}{W} \sum_{t=(j-1)W+1}^{jW} a^{l(i,t)} \quad (14)$$

$$p^{l(i,j)} = \max_{(j-1)W+1 \leq t \leq jW} \{a^{l(i,t)}\} \quad (15)$$

In this paper, the max-pooling function is applied and it outputs the maximal values of rectangular regions of its input.

In a fully connected layer, neurons between two adjacent layers are fully pairwise connected but neurons within the same layer share no connections. Then the SoftMax function is commonly adopted for classification tasks in the output layer. It is calculated as:

$$y_r(x) = P(c_r|x, \theta) = \frac{\exp(a_r(x))}{\sum_{j=1}^k \exp(a_j(x))} \quad (16)$$

The loss function can use mean squared error function and the cross-entropy function. In this paper, we used the cross-entropy function, which is given by:

$$E = -\sum_{i=1}^N \sum_{j=1}^k t_{ij} \ln(y_{ij}) \quad (17)$$

where t_{ij} is the indicator that the i th example belongs to the j th class, and y_{ij} is the output for example i , which is the value from the SoftMax function.

4. A novel fault diagnosis method

When the rolling bearings are in operation, the bearings with different types and different degrees of failure will generate vibrations which are typical non-stationary and non-linear multiple component vibration signals. The vibration signal collected by the acceleration sensor maybe the superposition of multiple vibration signals which include the interference signals. In the process of signal analysis and fault diagnosis, the interference signal makes it more difficult to extract valuable information. The VMD as an adaptive and non-recursive signal decomposition method that can reveal the weak transient impulse from complex vibration signals which has the advantages of effectively reducing pseudo components and modal aliasing. The 1-D CNN as a typical deep learning method that provides a structure in which both feature extraction and prediction are performed together in a single block, like genetic programming but unlike other traditional methods. 1-D CNN is superior to 2-D CNN in terms of computation burden and similar performance with relatively shallower architectures. This paper combines the above methods for fault diagnosis. The flowchart of the fault diagnosis method is shown in Figure 3.

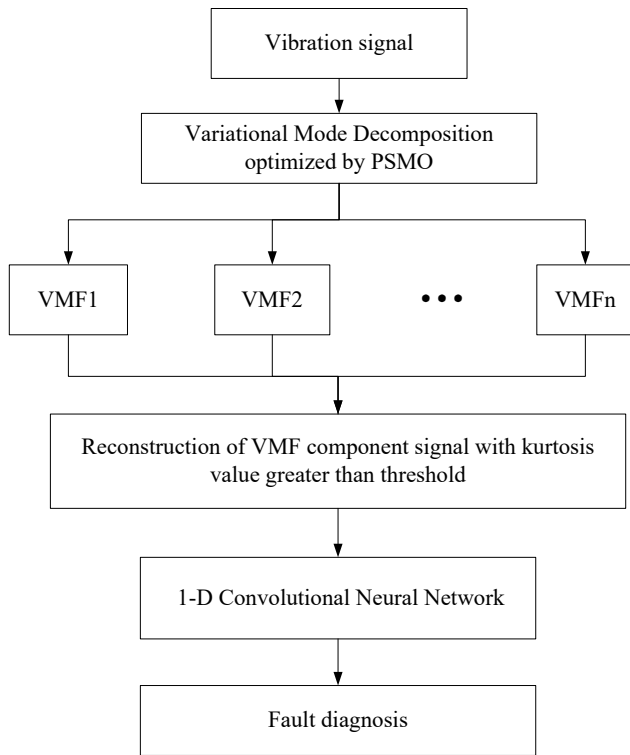


Figure 3. Flowchart of fault diagnosis method of rolling bearings

The specific steps of the proposed method are described as follows.

Step 1. The vibration signal data set is divided into training examples and testing examples.

Step 2. The modal number and the penalty factor of VMD are optimized by PSMO

Step 3. Training examples and testing examples are decomposed by optimized VMD, and some sets of IMF components are obtained.

Step 4. Signals are reconstructed by the VMF components obtained in the step 3.

Step 5. The 1-D CNN is trained using the reconstructed training examples.

Step 6. The effectiveness of the proposed fault diagnosis model is evaluated by experimental data of rolling bearings.

5. Verification and analysis

To verify the effectiveness of the proposed method in bearing fault diagnosis, two open-source datasets are used for research in this paper, including Case Western Reserve University (CWRU) Bearing Data Center dataset [41] and the Society for Mechanical Failure Prevention Technology (MFPT) dataset [42].

5.1 CWRU data

5.1.1 Experimental data and parameter settings.

The CWRU data is the experimental data of rolling bearings from the electrical engineering laboratory of Case Western Reserve University in this experiment [41]. The bearing model is a drive-end bearing (6205-2RSJEM SKF, deep groove ball bearing). As shown in Figure 4, the test bed consists of a 2-hp motor (left), a torque transducer and encoder (center), a dynamometer (right), and control electronics (not shown). The dynamometer is controlled so that desired torque load levels can be achieved. The test bearings support the motor shaft. Single point faults were introduced to the test bearings using electrical-discharge machining with fault diameters of 1.778×10^{-4} m, 3.556×10^{-4} m, 5.334×10^{-4} m, and the fault depth is 2.794×10^{-4} m.

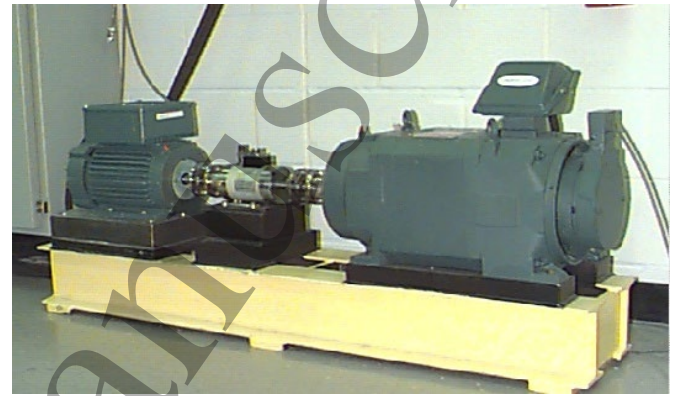


Figure 4. Rolling bearing test rig [41]

Table 1. Bearing health conditions

Bearing health condition	Fault diameter (mm)	Data label	Symbol
Rolling element fault	0.1778	0	B0
	0.3556	1	B1
	0.5334	2	B2
Inner race fault	0.1778	3	I0
	0.3556	4	I1
	0.5334	5	I2
Outer race fault	0.1778	6	O1
	0.3556	7	O2
	0.5334	8	O3
Normal	—	9	N

Vibration signals were collected using a 16-channel DAT recorder with sample rate of 12kHz. Each data set is made up of 1.2×10^5 points. The experimental rotating frequency is about 30Hz. The experimental data (Table 1), were collected vibration signals for normal and 3 different fault types (ball fault, inner race fault and outer race fault) with 3 different degrees (the different fault diameters correspond to the different fault degrees). Each of the ten conditions in the data sets is composed of 600 samples and the sample length of each signal is 4,096.

5.1.2 Vibration signal denoising

1) Parameters optimization of VMD based on PSMO.

In order to eliminate the noise of vibration signals of bearings, VMD method is used to decompose the signals. To obtain appropriate mode components, the parameters such as the modal number and the penalty parameter of VMD need to be optimized. We propose a novel PSMO method which has the advantages of global optimum of GA and the convergence speed of PSO. The parameters of PSMO are set as shown in Table 2. G_{max} is the maximum evolution algebra, M is the population size, c1 and c2 are learning factors, ω represents inertia weight, and q is the mutation probability.

Table 2. Parameters of PSMO

G_{max}	M	c1	c2	ω	q
10	10	1.5	1.5	1	0.2

The WSDA is taken as the fitness function. The smaller the WSDA value is, the better is the decomposition result. There are two termination conditions which are the number of iterations t and the WSDA value ε . As long as either $\varepsilon \leq 1e^{-7}$ or t = 10, the optimization result is output. In this paper, we select the data with 0.1778 inner race damage, which has 4096 sample length. In the fourth iteration, the minimum average difference is 0.0000425, and the corresponding optimal parameter combination is K = 8 and $\alpha = 6300$.

2) Signal decomposition. The vibration signals of 10 states are decomposed via the VMD method with optimized parameters obtained from 4.2.1. In the limited space, we introduce only the signal decomposition and reconstruction process of inner ring signal with a damage diameter of 0.14 feet. Figure 5 shows the time domain and frequency domain diagram of the original signal of inner ring fault. Figure 6 is the decomposition of the inner race fault signal.

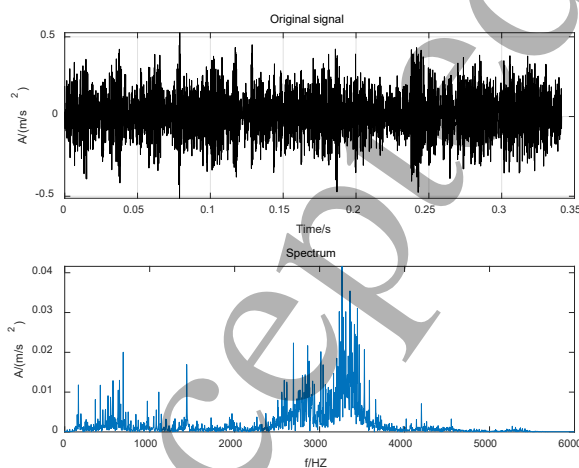


Figure 5. An original signal of inner race fault

3) Signal reconstruction. Signals are reconstructed according to the maximum kurtosis criterion, and the IMF component whose kurtosis is greater than the average kurtosis is selected for signal reconstruction. Figure 7 shows the time domain and frequency domain diagram of the reconstructed fault signal of the inner ring with a damage diameter of 0.14 feet. Compared with the Figure 4, it can be seen that the reconstructed signal removes the disorder components in the original signal and removes some noise signals. For other samples, there are similar differences between the other original signals and the reconstructed signals.

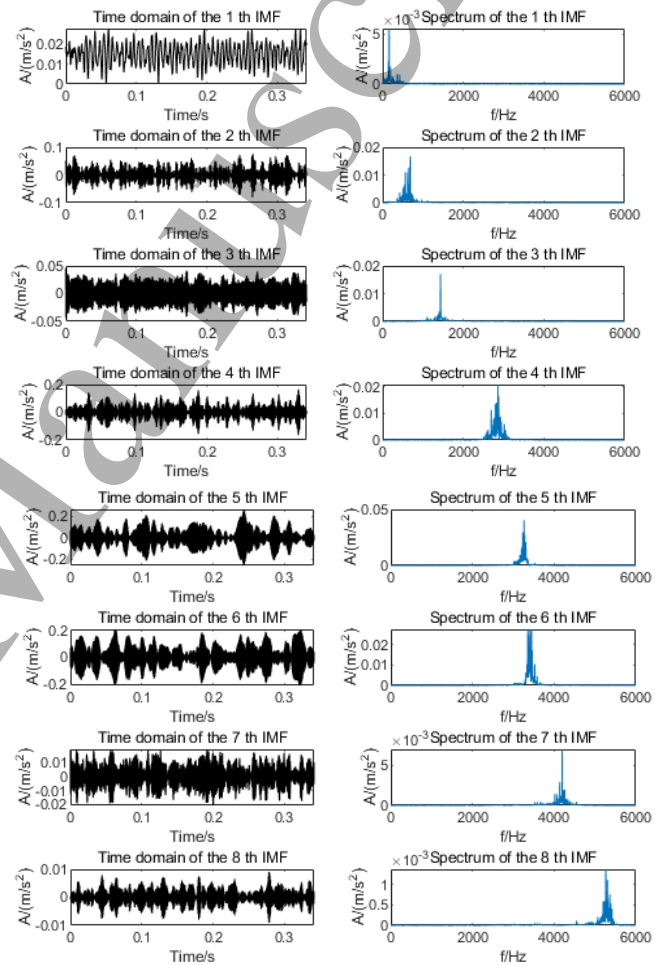


Figure 6. Decomposition of inner race fault signal

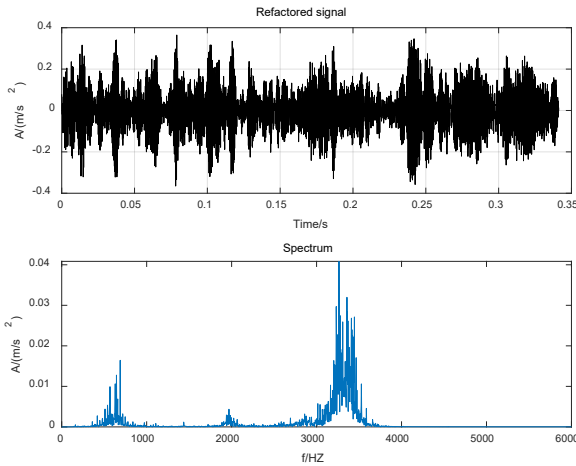


Figure 7. The reconstructed signal

To illustrate the effectiveness of the VMD, we adopt the index of permutation entropy (PE) to compare the original signal with the reconstructed signal. The parameters of PE need to be set; the embedding time delay is set as 5 and the embedding dimension is set as 4.

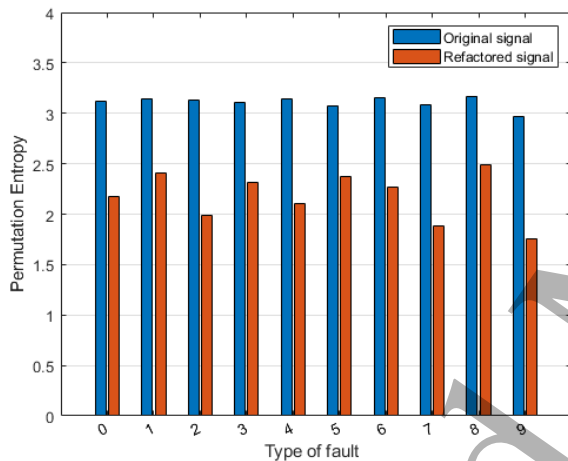


Figure 8. PE comparison of original signal and reconstructed signal

The smaller the value of PE is, the more regular and more deterministic the time series is. Contrarily, the greater the value of PE is, the more noisy and random the time series is. From Figure 8, it can be seen that the permutation entropy value of the reconstructed signal is smaller than that of the original signal. It indicates that the reconstructed signal has more information of working states and the original signal has more noise and chaos.

5.1.3 Fault diagnosis based on 1-D CNN

1) Parameter setting of 1-D CNN Reconstructed signals with 4096 data points are used as the input of 1-D CNN. Therefore, the size of input bearing signals is [4096 x 1]. The 1-D CNN model has a total of seven layers, including five convolutional and pooling layers, a fully connected layer and

a Softmax layer. The five layers of convolutional and pooling layers and their parameters are set as shown in Figure 9. Relu function [43] as the activation function is used in all layers except the Softmax layer. In Table 3, “same” means that the output feature map has the same spatial dimensions as the input feature map. Zero padding is introduced to make the shapes match as needed, equally on every side of the input map. Here, “same” tries to pad evenly left and right, but if the number of columns to be added is odd, it will add the extra column to the right. Here, “valid” means no padding and only ever drops the right-most columns (or bottom-most rows).

Table 3. Layers of 1-D CNN

layer	convolution				pooling		
	filter	size	stride	padding	size	stride	padding
1	16	64	16	same	2	2	valid
2	32	3	1	same	2	2	valid
3	64	3	1	same	2	2	valid
4	64	3	1	same	2	2	valid
5	64	3	1	valid	2	2	valid

As the Adam optimization algorithm (Adam) may reduce the oscillations along the path of the steepest descent towards the optimum that is sometimes caused by stochastic gradient descent algorithm [44], we use the Adam algorithm to update the parameters of the deep NN. The stochastic gradient descent with momentum update is

$$\omega_t = \omega_{t-1} - \alpha * \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (18)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (19)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (20)$$

where m_t and v_t are the first-order moment estimation and second-order moment estimation of the gradient, respectively, \hat{m}_t and \hat{v}_t is the correction of m_t and v_t , which can be approximated to the unbiased estimation of expectation. The parameter t represents the number of times, β_1 and β_2 are constants, controlling exponential attenuation. Also, α is called the learning rate or step size factor, which controls the update ratio of weights, ϵ is a very small number in order to prevent division by zero in the implementation. Here, we set the momentum β_1 at 0.9 and β_2 at 0.99, the learning rate α at 0.001, ϵ at 10e-8 and the maximum number of epochs to use for training at 10.

2) Diagnosis results and analysis.

The 5000 samples are taken as the training set and 1000 samples are taken as testing set in the experiment. The original signal and reconstructed signal are used respectively as the input layer of 1-D CNN to verify the diagnosis accuracy of the proposed method. The average diagnosis accuracy for running 10 times was 99.6% for reconstruction signals and 98.7% for original signals. Table 4 records the diagnosis accuracy for original signals and for reconstructed signals at each running time. From the diagnosis accuracy, we can see that at every

running time, the proposed method can obtain a higher diagnosis accuracy which indicates that the VMD method is very important in the process of fault diagnosis to obtain higher diagnosis accuracy.

Table 4. Accuracy of diagnosis of vibration signals

Run	Accuracy of diagnosis (%)	
	original signals	Reconstructed signals
1	98.1	99.7
2	99.2	99.8
3	98.0	99.8
4	98.8	99.7
5	99.1	99.6
6	96.0	99.7
7	99.1	99.9
8	99.0	99.5
9	99.0	99.4
10	98.3	99.8

In the limited space, we only provide a confusion matrix of the classification results for each condition with testing data at the 4th run which is shown in Figure 9. The reason for selecting the results of the 4th run is that the diagnosis accuracy of the 4th run is near to the average accuracy. In Figure 9, 1-D CNN misclassified 1% of the testing examples of B1 condition as N condition, and misclassified 10% of the testing examples of B2 condition as B1 condition for the original signals. Our method misclassified 3% of the testing examples of B2 condition as B1 condition for the constructed signals.

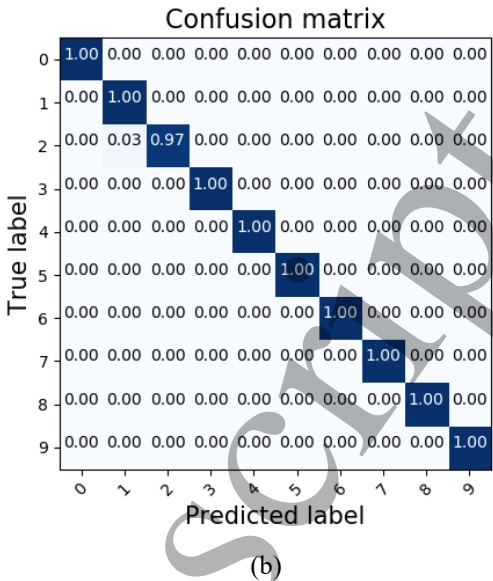
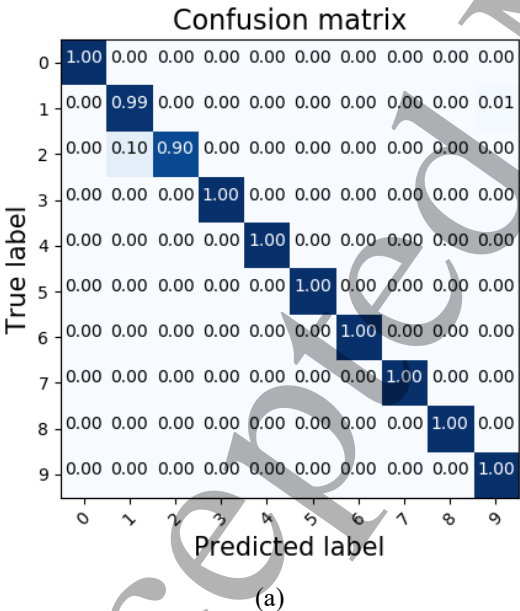


Figure 9. Confusion matrix of original signals (a) and reconstructed signals (b)

In order to describe the overall perspective of confusion matrix under 10 runs in limited space, the numbers of misclassifications in the confusion matrix under 10 runs are count in the Table 5 and Table 6 respectively for original signals and for reconstructed signals.

As can be seen from Table 5 that out of the 10 runs, the condition of B1 misclassified as B0 one time; and the condition of B2 is misclassified as B0 seven times, as B1 one times, and as O1 one time; the condition of I1 is misclassified as B0 two times, as B2 one times, as I0 one time and O1 three times. By contrast, the content of Table 6 is simpler than Table 5. Only the conditions of B2 and I0 occurred misclassification. The condition of B2 is misclassified as B0 six times and B1 three times. The condition of I0 is misclassified as I1 three times.

Table 5. Number of misclassifications for original signals

		Prediction									
		B0	B1	B2	I0	I1	I2	O0	O1	O2	N
True	B0										
	B1	1									
	B2	7	1						1		
	I0										
	I1	2		1	1				3		
	I2										
	O0										
	O1										
	O2										
	N										

Table 6. Number of misclassifications for reconstructed signals

		Prediction									
		B0	B1	B2	I0	I1	I2	O0	O1	O2	N
True	B0										
	B1										
	B2	6	3								
	I0					3					
	I1										
	I2										
	O0										
	O1										
	O2										
	N										

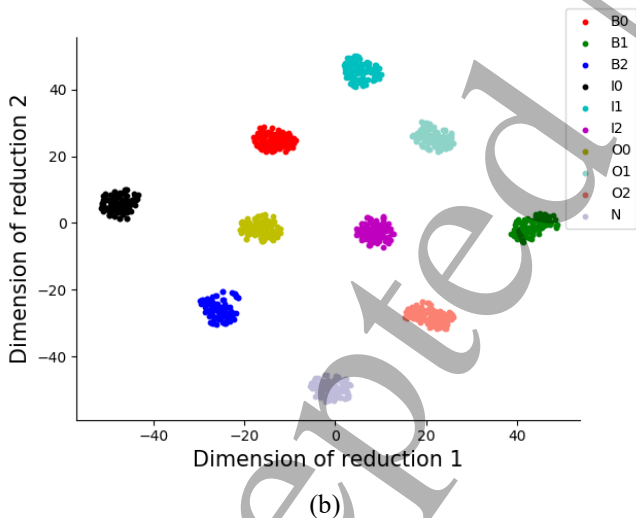
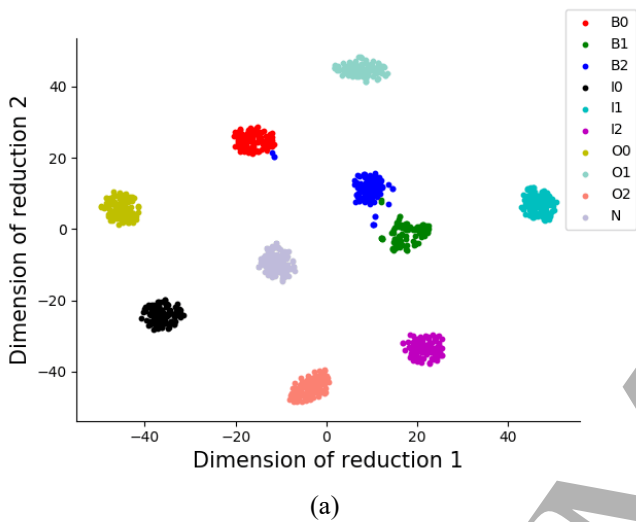


Figure 10. Data distribution at full connection layer for the original signals (a) and the reconstructed signals (b)

It is known that the classification accuracy is influenced by the distribution of samples. To observe more clearly whether there is the difference of distribution between the original signals and the constructed signals, we use the t-SNE [45]

dimension reduction algorithm to visualize the distribution of high dimensional datasets at the full connection layer.

It can be seen from Figure 10 that, for the original signals, there are some overlaps among 3 different types of datasets, which are the conditions of B0, B1, and B2. It is more easily confused among these 3 different fault degrees of rolling element fault type in the process of diagnosis. It is consistent with Figure 9 that the conditions of B0, B1 and B2 are easily misclassified. It is obvious that for the reconstructed signals, there are clear boundaries between different types of data sets. This partly explains why the diagnosis accuracy of the reconstructed signal is higher than that of the original signal.

5.1.4 Comparison

To verify further the effectiveness of the proposed method, we use Long short-term memory (LSTM), Random Forest (RF), and Support Vector Machine (SVM) to classify the faults of the original data and the faults of the reconstructed data. The parameters are set as follows: for LSTM, the layers is set to two, and the neurons numbers of the two layers are set to 16 and 32 respectively; for RF, the number of trees in the forest is set to 100. Information Entropy is used to measure the performance of splitting quality, the maximum depth of the random tree is set to none mode, the meaning nodes are expanded until all leaves are pure. For SVM, radial basis function is introduced as the kernel function of support vector machine. The above three methods have been used for 10 runs each to classify the original data and classify the reconstructed data.

Table 7. Diagnosis accuracy of three methods

Run	Accuracy of diagnosis (%)					
	Reconstructed signals			original signals		
	LSTM	RF	SVM	LSTM	RF	SVM
1	93.1	89.5	85.8	91.6	82.7	73.8
2	94.3	90.7	79.9	94.3	84.3	74.3
3	95.2	91.2	79.4	92.2	83.1	77.2
4	93.6	88.3	82.3	89.9	82.3	76.5
5	93.9	87.5	83.5	90.2	86.5	80.9
6	95.5	89.2	85.6	93.3	83.6	75.6
7	96.4	92.3	78.5	85.6	85.2	78.8
8	92.3	86.4	86.2	89.7	82.1	76.6
9	93.4	90.5	82.3	91.2	83.3	77.7
10	95.2	89.7	83.8	90.3	81.4	73.4

Table 7 contains the results of the three methods. Compared with Table 4, we can see that the proposed method is better than other methods, both on the original data and the reconstructed data. We also conclude that the diagnosis accuracy using the reconstructed data is better than the same with the original data, no matter whether 1-D CNN, LSTM, RF or SVM is used. It demonstrates that the optimized VMD is essential to improve the diagnosis accuracy.

5.2 MFPT data

5.2.1 Experimental data and parameter settings

The used data is the experimental data of rolling bearings from the Machinery Failure Prevention Technology (MFPT) Association in this experiment, which comprised two real-world kinds of fault states and one kind of normal state [42]. The motor speed was 25 Hz. Baseline data were gathered at a sampling frequency of 97,656 Hz and under a load of 270 lbs. Outer race fault data were gathered at a sampling frequency of 48,828 Hz and under seven different loads (25, 50, 100, 150, 200, 250, 300 lbs), and inner race fault data were gathered at a sampling frequency of 48,828 Hz and under seven different loads (0, 50, 100, 150, 200, 250, and 300 lbs). In this paper, five conditions of bearing data are used for fault diagnosis, shown in Table 8. Each sample is a collected vibration signal segment consisting of 4,096 sampling data points and 500 samples is included in each condition of bearing data.

Table 8. Bearing health conditions

Bearing health condition	Load (lbs)	Data label	Symbol
Normal	270	0	N
Inner race fault	0	1	I1
	50	2	I2
Outer race fault	25	3	O1
	50	4	O2

5.2.2 Vibration signal denoising

To eliminate the noise of vibration signals of bearings, firstly we optimize the modal number and the penalty parameter of VMD by PSMO, then decompose signals by VMD, and finally reconstruct signals of bearing data.

1) *Parameters optimization of VMD based on PSMO.* To obtain appropriate mode components, the parameters such as the modal number and the penalty parameter of VMD need to be optimized by PSMO. The parameters of PSMO are set as shown in Table 9. G_{max} is the maximum evolution algebra, M is the population size, $c1$ and $c2$ are learning factors, ω represents inertia weight, and q is the mutation probability.

Table 9. Parameters of PSMO

G_{max}	M	$c1$	$c2$	ω	q
10	10	1.5	1.5	1	0.2

The WSDA is taken as the fitness function. The smaller the WSDA value is, the better is the decomposition result. There are two termination conditions which are the number of iterations t and the WSDA value ε . As long as either $\varepsilon \leq 1e^{-7}$ or $t = 10$, the optimization result is output. In this paper, we select the data with 0lbs inner race damage, which has 4,096 sample length. In the fourth iteration, the minimum average difference is 0.0000994, and the corresponding optimal parameter combination is $K = 6$ and $\alpha = 3623$.

2) *Signal decomposition.* The vibration signals of 5 states are decomposed via the VMD method with $K = 6$ and $\alpha = 3,623$. In the limited space, we introduce only the signal decomposition and reconstruction process of inner ring signal with load of 0 lbs load. Figure 11 shows the time domain and frequency domain diagram of the original signal of inner ring fault. Figure 12 is the decomposition of the inner race fault signal.

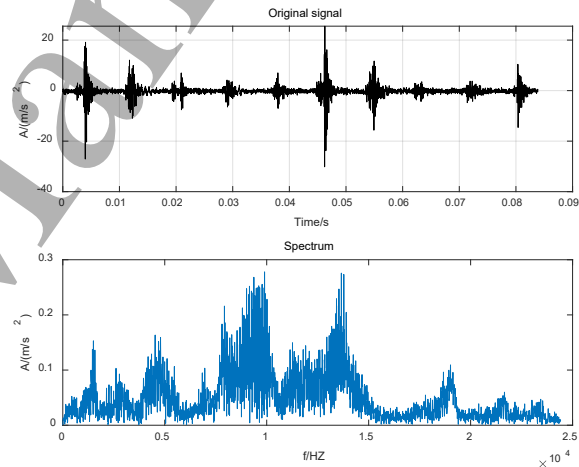


Figure 11. An original signal of inner race fault

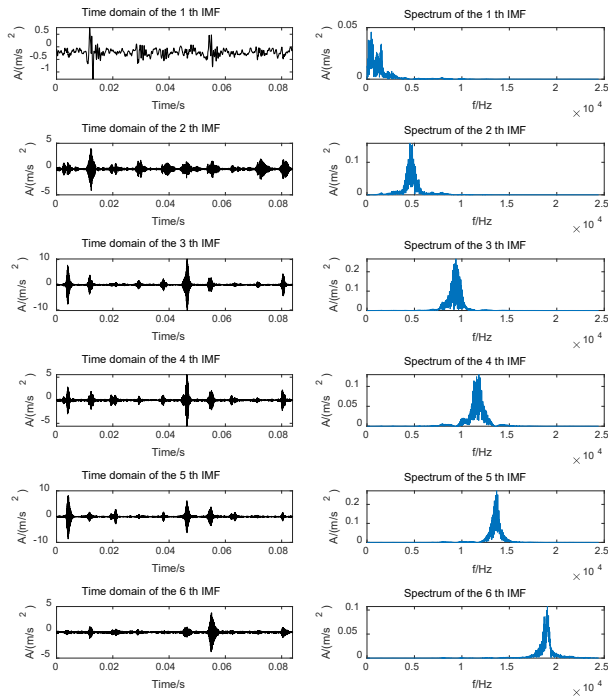


Figure 12. Decomposition of inner race fault signal

3) *Signal reconstruction.* Signals are reconstructed according to the maximum kurtosis criterion, and the IMF component whose kurtosis is greater than the average kurtosis is selected for signal reconstruction. Figure 13 shows the time domain and frequency domain diagram of the reconstructed fault signal of the inner ring with a damage 0 lbs load.

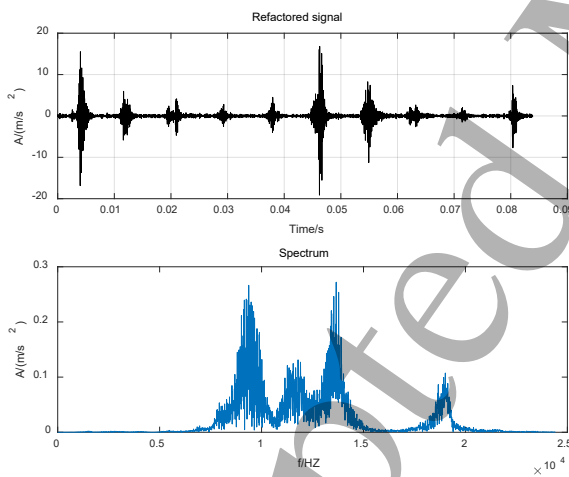


Figure 13. The reconstructed signal

Compared with the Figure 11, it can be seen that the reconstructed signal removes the disorder components in the original signal and removes some noise signals. For other samples, there are similar differences between the other original signals and the reconstructed signals.

To illustrate the effectiveness of the VMD, we adopt the PE index to compare the original signal with the reconstructed

signal. The embedding time delay and the embedding dimension of PE are set to 5 and 4 respectively.

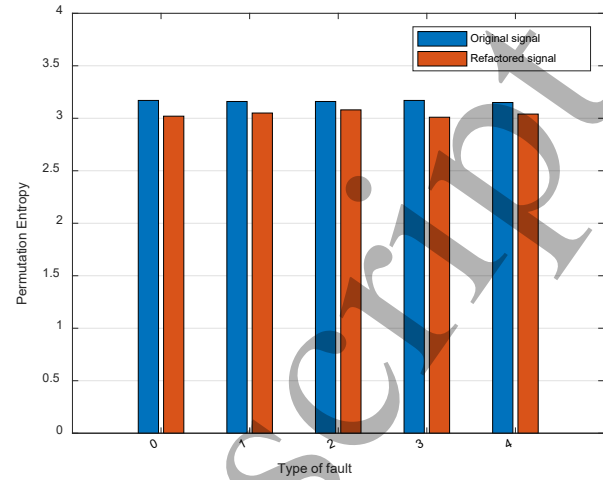


Figure 14. PE comparison of original signal and reconstructed signal

From Figure 14, it can be seen that the PE value of the reconstructed signal is smaller than that of the original signal. It indicates that the reconstructed signal has more information of working states compared with the original signal.

5.2.3 Fault diagnosis based on 1-D CNN

1) *Parameter setting of 1-D CNN.* Reconstructed signals with 4,096 data points are used as the input of 1-D CNN. Therefore, the size of input bearing signals is [4,096 x 1]. The 1-D CNN model has a total of seven layers, including five convolutional and pooling layers, a fully connected layer and a Softmax layer. The five layers of convolutional and pooling layers and their parameters are set as shown in Table 10. Relu function [42] is used as the activation function in all layers except the Softmax layer and the Adam algorithm [43] is used to update the parameters of the deep NN.

Table 10. Layers of 1-D CNN

layer	convolution				pooling		
	filter	size	stride	padding	size	stride	padding
1	16	64	16	same	2	2	valid
2	32	3	1	same	2	2	valid
3	64	3	1	same	2	2	valid
4	64	3	1	same	2	2	valid
5	64	3	1	valid	2	2	valid

2) Diagnosis results and analysis.

The 2,000 samples are taken as the training set and 500 samples are taken as testing set in the experiment. The original signal and reconstructed signal are used respectively as the input layer of 1-D CNN to verify the diagnosis accuracy of the proposed method. The average diagnosis accuracy for running 10 times was 96.1% for reconstruction signals and 91.2% for original signals. Table 11 records the diagnosis accuracy for

original signals and for reconstructed signals at each running time. From the diagnosis accuracy, we can see that at every running time, the proposed method can obtain a higher diagnosis accuracy which indicates that the VMD method is very important in the process of fault diagnosis to obtain higher diagnosis accuracy.

Table 11. Accuracy of diagnosis of vibration signals

Run	Accuracy of diagnosis (%)	
	original signals	Reconstructed signals
1	88.2	97.6
2	93.5	97.2
3	90.8	95.9
4	89.9	95.5
5	91.3	93.9
6	90.7	96.3
7	91.6	97.7
8	93.1	95.8
9	92.1	96.3
10	90.6	94.5

Figure 15 provides a confusion matrix of the classification results for each condition with testing data at the 6th run. The reason for selecting the results of the 6th run is that the diagnosis accuracy of the 4th run is near to the average accuracy.

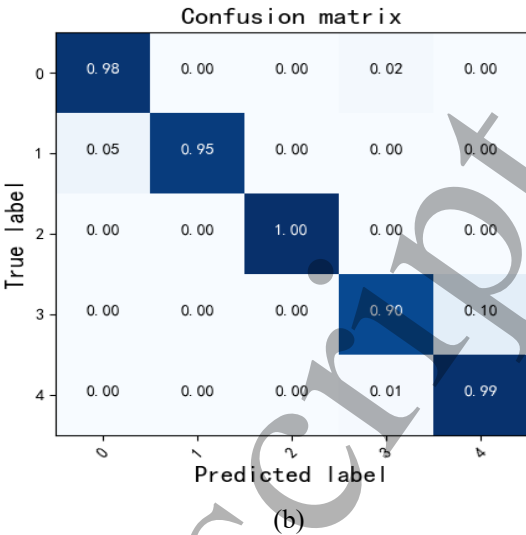
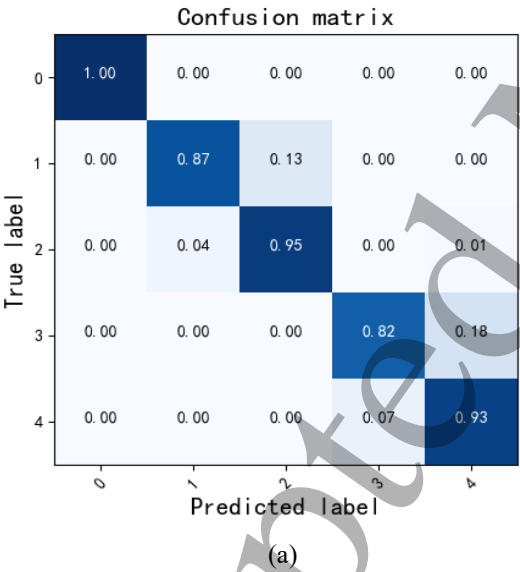
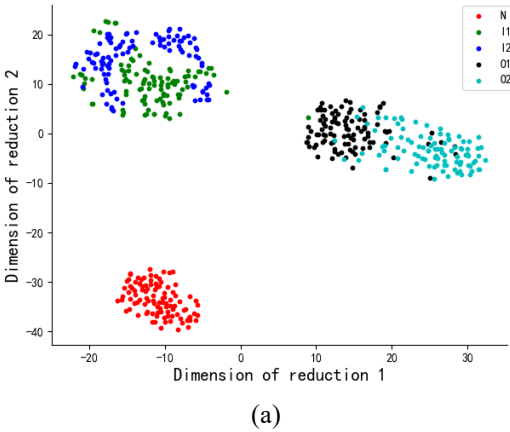


Figure 15. Confusion matrix of original signals (a) and of reconstructed signals (b)

In Figure 15, for the original data, 1-D CNN misclassified 13% of the testing examples of I1 condition as I2 condition, and misclassified 18% of the testing examples of O1 condition as O2 condition for the original signals; For the reconstructed data, 1-D CNN misclassified 5% of the testing examples of I2 condition as N condition, and misclassified 10% of the testing examples of O1 condition as O2 condition.

It is well known that the classification performance is influenced by the distribution of samples. In order to observe more clearly whether there is the difference of distribution between the original signals and the constructed signals, we use the t-SNE [44] dimension reduction algorithm to visualize the data distribution of high dimensional datasets at the full connection layer.



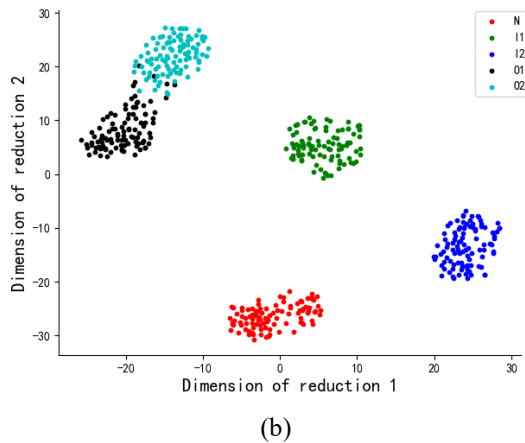


Figure 16. Data distribution at full connection layer for the original signals (a) and the reconstructed signals (b)

It can be seen from Figure 16 that, for the original signals, there are some overlap between I1 and I2 and some overlap between O1 and O2, which is consistent with the Figure 15. For the reconstructed signals, there are nearly clear boundaries between different types of data sets. To some extent, this explains why the diagnosis accuracy of the reconstructed signal is higher than that of the original signal.

5.2.4 Comparison

To verify further the effectiveness of the proposed method, we use Long short-term memory (LSTM), Random Forest (RF), and Support Vector Machine (SVM) to classify the faults of the original data and the faults of the reconstructed data, then we compare them with the results of our proposed method.

The parameters are set as follows: for LSTM, the layers is set to two, and the neurons numbers of the two layers are set to 16 and 32 respectively; for RF, the number of trees in the forest is set to 100. Information Entropy is used to measure the performance of splitting quality, the maximum depth of the random tree is set to none mode, the meaning nodes are expanded until all leaves are pure. For SVM, radial basis function is introduced as the kernel function of support vector machine. The above three methods have been used for 10 runs on the original data and on the reconstructed data. The results are shown in Table 12.

Table 12. Diagnosis accuracy of three methods

Run	Accuracy of diagnosis (%)					
	Reconstructed signals			original signals		
	LSTM	RF	SVM	LSTM	RF	SVM
1	88.8	71.2	66.7	80.2	58.7	64.3
2	86.7	73.6	70.5	79.8	59.6	66.2
3	91.3	72.3	71.2	81.3	60.2	63.5
4	92.6	70.9	67.4	80.5	61.5	62.9
5	91.9	74.5	66.1	81.2	62.3	64.1
6	87.6	76.3	72.2	80.5	56.9	63.9
7	85.3	75.6	66.8	79.6	58.3	62.3
8	88.1	71.5	71	82.2	61.3	63.4
9	89.4	74.2	69.8	81.4	62.2	65.2
10	87.5	73.5	71.2	80.3	61.4	63.3

Compared with Table 11, we can see that the proposed method is better than other methods, both on the original data and the reconstructed data. We also conclude that the diagnosis accuracy using the reconstructed data is better than the same with the original data, no matter whether 1-D CNN, LSTM, RF or SVM is used. It demonstrates that the optimized VMD is essential to improve the diagnosis accuracy.

6. Conclusions

In this paper, to address two challenging issues of how to effectively eliminate noise and extract valid fault features from the vibration signals for bearing fault diagnosis, a novel hybrid fault diagnosis method, PSMO, based on the optimized VMD and 1-D CNN has been proposed to realize fault diagnosis of rolling bearings.

PSMO, which has the advantages of strong global search ability and fast convergence speed by combining PSO and GA, and WSDA as the fitness function are proposed to realize the optimization of K and α in VMD. From the index of permutation entropy (PE), the permutation entropy value of the reconstructed signal is smaller than that of the original signal, which indicates that the reconstructed signal has more information of working states than the original signal.

1-D CNN which fuses feature extraction and classification operations into a single machine learning body to optimize jointly the classification performances, has been adopted to realize fault diagnosis of the reconstructed vibration signals. For the CWRU data and MFPT data, our method has better

classification accuracy of 99.6% for CWRU data and 96.1% for MFPT data, compared with LSTM, RF, and SVM. Meanwhile, VMD denoising signals by PSMO is very important in the process of fault diagnosis, because the reconstructed data has been better adopted for this task than the original data.

Acknowledgements

This research was funded by the National Natural Science Foundation of China, No. 51105291 and No. 51605364, by the Shaanxi Provincial Science and Technology Department, No. 2020GY-124, NO.2019GY-125 and No.2018JQ5127, and by the Key Laboratory Project of Department of Education of Shaanxi Province, No.19JS034 and No. 18JS045. Also, this work is supported by Brunel University London (UK) and the National Fund for Study Abroad (China).

References

- [1] Lei Y G, Jia F, Zhou X and Lin J 2015 A deep learning-based method for machinery health monitoring with big data *J. Mech. Eng.* **51** 49 – 56
- [2] Zhang L and Nandi A K 2007 Fault classification using genetic programming *Mech. Syst. Signal Process.* **21** 1273-1284
- [3] Wong M L D, Jack L B and Nandi A K 2006 Modified self-organising map for automated novelty detection applied to vibration signal monitoring *Mech. Syst. Signal Process.* **20** 593-610
- [4] Guo H, Jack L B and Nandi A K 2005 Feature generation using genetic programming with application to fault classification *IEEE Trans. Syst. Man. Cybern. Part B.* **35** 89-99
- [5] Jack L B and Nandi A K 2002 Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms *Mech. Syst. Signal Process.* **16** 373-390
- [6] Jack L B and Nandi A K 2001 Support vector machines for detection and characterization of rolling element bearing faults *Proc. Instn. Mech. Engrs. Part C.* **215** 1065-1074
- [7] Seera M, Wong M L D and Nandi 2017 Classification of ball bearing faults using a hybrid intelligent model *Applied Soft Computing.* **57** 427-435
- [8] Ahmed H O A and Nandi A K 2018 Three-stage hybrid fault diagnosis for rolling bearings with compressively sampled data and subspace learning techniques *IEEE Trans. Ind Electron.* **66** 5516-5524
- [9] Ahmed H and Nandi A K 2019 Compressive sampling and feature ranking framework for bearing fault classification with vibration signals *IEEE Access.* **6** 44731-44746
- [10] Ahmed H O A, Wong M L D and Nandi A K 2018 Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features *Mech. Syst. Signal Process.* **99** 459-477
- [11] Ahmed H and Nandi A K 2020 Condition monitoring with vibration signals: compressive sampling and learning algorithms for rotating machines *Published by John Wiley & Sons, Chichester, West Sussex, UK, 2020 (ISBN 978-1-119-54462-3).*
- [12] Xu L, Chatterton S and Pennacchi P 2020 Rolling element bearing diagnosis based on singular value decomposition and composite squared envelope spectrum *Mech. Syst. Signal Process.* **148** 107174
- [13] Zou L Q, Chen G J, Xing J J and Jiang C H 2014 Fault diagnosis method based on LMD sample entropy and SVM for reciprocating compressors *Noise and Vibration Control.* **34** 174-177
- [14] Rilling G, Flandrin P and Goncalves P 2003 On empirical mode decomposition and its algorithms *IEEE-EURASIP Workshop Nonlinear Signal Image Processing (NSIP).* **3** 8-11
- [15] Guo W and Tse P W 2010 Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis *Proc. IEEE Conf. Prognostics Health Management PHM'10* pp 1-7
- [16] Lei Y G and Zuo M J 2009 Fault diagnosis of rotating machinery using an improved HHT based on EEMD and sensitive IMFs *Mech. Sci. Technol.* **20** 125701
- [17] Li K, Su L, Wu J, Wang H and Peng C 2017 A rolling bearing fault diagnosis method based on variational mode decomposition and an improved kernel extreme learning machine *J. Applied Sci.* **7** 1004
- [18] Dragomiretskiy K and Zosso D 2014 Variational mode decomposition *IEEE Trans. Signal Process.* **62** 531-544
- [19] Wang X and Yan W 2017 Fault diagnosis of roller bearings based on the variational mode decomposition and SVM *Shock and Vibration.* **36** 252-256
- [20] Gu J, Peng Y, Lu H, Cao S and Cao B 2020 Fault diagnosis of spindle device in hoist using variational mode decomposition and statistical features *Shock and Vibration* **6** 1-14
- [21] Liu J C 2019 Rolling bearing fault diagnosis based on parameter optimization VMD and sample entropy *Acta Automatica Sinica.* **190345**
- [22] Upadhyay A and Pachori R B 2015 Instantaneous voiced/non-voiced detection in speech signals based on variational mode decomposition *J. Franklin Inst.* **352** 2679-2707
- [23] Aneesh C, Kumar S, Hisham P M and Soman K P 2015 Performance comparison of variational mode decomposition over empirical wavelet transform for the classification of power quality disturbances using support vector machine *Proc. Comput. Sci.* **46** 372-380
- [24] Jie B. Fault diagnosis of bearing combining parameter optimized variational mode decomposition based on genetic algorithm with 1.5-dimensional spectrum *J. Propulsion Tech.* **38** 1618-1624
- [25] Zhang Y and Wang A 2020 Research on the fault diagnosis method for rolling bearings based on improved VMD and automatic IMF acquisition *Shock and Vibration.* **1-19**
- [26] Zhan Z H and Zhang J 2009 Adaptive particle swarm optimization *IEEE Trans. Syst. Man. Cybern. B. Cybern.* **39** 2679-2707
- [27] Ho S Y, Lin H S, Liauh W H and Ho S J 2008 OPSO: Orthogonal particle swarm optimization and its application to task assignment problems *IEEE Trans. Syst. Man. Cybern. A. Syst. Hum.* **38** 288-298
- [28] Kiranyaz S, Gastli A, Ben-Brahim L, Alemadi N and Gabbouj M 2018 Real-time fault detection and identification for MMC using 1D convolutional neural networks *IEEE Trans. Ind Electron.* **99** 1-1

- [29] Lei Y, Yang B, Jiang X, Jia F, Li N and Nandi A K 2020 Application of machine learning to machine fault diagnosis: A review and roadmap *Mech. Syst. Signal Process.* **138** 106587
- [30] Dong S J, Pei X W, Wu W L, Tang B P and Zhao X X 2020 Rolling bearing fault diagnosis method based on multilayer noise reduction technology and improved convolutional neural network *J. Mech. eng.* 1-9
- [31] Wang Q H, Nandi A K, Yu Y, Darwish M and Ahmed H O 2020 Fault detection and classification in MMC-HVDC systems using learning methods *Sensors.* **20** 20164438
- [32] Li H, Zhang Q, Qin X R and Sun Y 2018 Bearing fault diagnosis method based on short time fourier transform and convolutional neural network *Vibration and Shock.* **37** 124 – 131
- [33] Huang S, Tang J, Dai J and Wang 2019 Signal status recognition based on 1DCNN and its feature extraction mechanism analysis *Sensors.* **19** 19092018
- [34] Wang H, Liu Z, Peng D and Qin Y 2019 Understanding and learning discriminant features based on multi-attention 1DCNN for wheelset bearing fault diagnosis *IEEE Trans. Ind. Inform.* **99** 1-1
- [35] Huang S Z, Tang J, Dai J Y, Wang Y Y and Dong J J 2020 1DCNN fault diagnosis based on cubic spline interpolation pooling *Shock and Vibration.* **20** 1949863
- [36] Zhao R, Yan R, Chen Z, Mao K and Gao R X 2019 Deep learning and its applications to machine health monitoring *Mech. Syst. Signal Process.* **115** 213-237.
- [37] Sumba J C, Quinde I R, Ochoa L E, JCT Martinez J C T and Morales-Menendez R 2019 Intelligent fault diagnosis for rotating machines using deep learning *Smart Sust. Manu. Syst.* **3** 28-40
- [38] Bandt C and Pompe B 2002 Permutation entropy: a natural complexity measure for time series *Physics Review Letters.* **88** 1-4
- [39] Yan R Q, Liu Y B, Robert X and Cao G 2012 Permutation entropy: A nonlinear statistical measure for status characterization of rotary machines *Mech. Syst. Signal Process.* **29** 474-484
- [40] Lecun Y, Bengio Y and Hinton G 2015 Deep learning *Nature.* **521** 436-444
- [41] K.A. Loparo. Bearing vibration data set. *Case Western Reserve University.* (<http://www.eecs.cwru.edu/laboratory/bearing/download.htm>).
- [42] E. Bechhoefer. (2016). A Quick Introduction to Bearing Envelope Analysis. MFPT Data. [Online]. Available: <http://www.mfpt.org/FaultData/Fault-Data.htm.Set>
- [43] Glorot X, Bordes A and Bengio Y 2011 Deep sparse rectifier neural networks *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS).* 315-323
- [44] Kingma D and Ba J 2014 Adam: A method for stochastic optimization *Comput. Sci.*
- [45] Van d M L and Hinton G 2008 Visualizing high-dimensional data using t-sne *Journal of Machine Learning Research.* **9** 2579-2605